

Practical Data Analytics with Python for Beginners

Phân Tích Dữ Liệu Thực tế với Python cho người mới học

Số hóa dữ liệu, trực quan hóa dữ liệu, phân tích dữ liệu là những nhu cầu rất lớn trong thời kì chuyển đổi số hiện nay. Những kiến thức trong khóa học này giúp bạn có một công việc vững chắc cho hiện tại và tương lai liên quan đến dữ liệu. Mục tiêu của khóa học là chuẩn bị những kỹ năng lập trình Python và phân tích dữ liệu nền tảng để đưa bạn từ một beginner hoặc dân trái ngành trở thành Chuyên viên Phân tích Dữ liệu (Data Analyst). Sau khóa học này, bạn có thể tự tin sử dụng Python cho công việc của mình một cách chủ động, sáng tạo, thậm chí vượt ra ngoài khuôn khổ những gì được học tại lớp, hoặc ứng tuyển vào các công ty chuyên về phân tích dữ liệu. Và đây cũng là điểm khác biệt giữa khóa đào tạo của chúng tôi với các khóa ngắn hạn khác, vốn thường tập trung vào kỹ năng sử dụng các công cụ có sẵn như Excel, Power BI, Tableau. Bám theo triết lý “*a picture is worth a thousand words*” (tạm dịch: *một hình ảnh đáng giá hàng ngàn từ ngữ*), chúng tôi thiết kế khóa học này với trọng tâm là các kỹ năng Trực quan hóa Dữ liệu (Data Visualization), giúp các bạn khả năng “biến” những con số hay ký tự khô khan trở thành những hình ảnh biết nói. Ngoài ra, khóa học cũng cung cấp các buổi hướng dẫn thực hành lập trình Python cơ bản một cách kỹ càng cho các học viên là dân trái ngành, chưa có kiến thức/kinh nghiệm lập trình, để hỗ trợ các học viên này nắm được kỹ năng lập trình thật vững chắc.

Đồng thời, đây cũng là khóa học đầu tiên trong một lộ trình từ cơ bản đến nâng cao trong Data Science, nhằm đào tạo ra các Data Scientist và Data Science Leader/Manager tại CyberLab. Một số khóa học tiếp theo trong lộ trình là Xác suất thống kê cho phân tích dữ liệu, Trí tuệ nhân tạo cho phân tích dữ liệu (bao gồm 2 khóa từ Machine Learning cơ bản đến Deep Learning nâng cao), và cuối cùng là Quản trị dự án phân tích dữ liệu. Chúng tôi cũng sẽ tiếp tục mở các lớp đào tạo nâng cao kỹ năng phân tích dữ liệu trong các lĩnh vực cụ thể với các khóa (ngắn) chuyên sâu, như: Phân tích dữ liệu Nhân sự, Phân tích dữ liệu Marketing, Phân tích dữ liệu Tài chính, Phân tích dữ liệu Logistics.

Mục tiêu (Learning Objectives):

1. Bạn sẽ được đào tạo để trở thành một Chuyên gia Phân tích Dữ liệu (Data Analyst) chuyên nghiệp
2. Phát triển các tố chất và kỹ năng cần có của một Chuyên gia Phân tích Dữ liệu/Nhà Khoa học Dữ liệu chuyên nghiệp
3. Hiểu rõ và Thực hiện được các bước cần tiến hành trong một dự án phân tích dữ liệu
4. Nắm vững và Thực hành kỹ thuật code Python, kỹ năng phân tích & giải quyết vấn đề phục vụ cho việc phân tích dữ liệu
5. Phát triển tư duy, logic nền tảng, cốt lõi trong nghề Phân tích Dữ liệu
6. Rèn luyện và phát huy kỹ năng làm việc nhóm trong Phân tích Dữ liệu
7. Có khả năng giải thích và truyền đạt kết quả phân tích dữ liệu một cách thuyết phục
8. Có khả năng xây dựng dự án phân tích dữ liệu thực tế
9. Nắm vững các công thức để tay nghề luôn được vững chắc, cập nhật công nghệ
10. Vừa chuyên nghiệp hóa chuyên môn, vừa thông thạo các kỹ năng mềm trong làm việc

Kết quả sau khóa học (Learning Outcomes):

1. Biết cách lập trình Python thay cho các công cụ truyền thống hoặc có sẵn (như Excel, Power BI, Tableau...) để thực hiện các dự án phân tích dữ liệu
2. Nắm vững kiến thức về phân tích dữ liệu với Matplotlib, Pandas, Dash, Plotly
3. Biết cách vận dụng các thư viện trong Python để thu thập, tiền xử lý, làm sạch dữ liệu

4. Biết cách vận dụng các thư viện khác nhau trong Python để thực hiện trực quan hóa (Visualize) dữ liệu và kết quả phân tích dữ liệu
5. Có khả năng thực hiện phân tích dữ liệu nâng cao, phân tích EDA với Cơ sở dữ liệu quan hệ và truy vấn SQL
6. Nắm vững và Thực hành kỹ năng xây dựng Dashboard có khả năng tương tác trên website
7. Được truyền kinh nghiệm "máu lửa" từ các Giảng viên đã kinh qua nhiều dự án trong và ngoài nước
8. Được đào tạo theo một LỘ TRÌNH chuẩn hóa qua các phương pháp giảng dạy TÍCH CỰC (Active Learning)
9. Có khả năng đáp ứng các nhu cầu tuyển dụng trong các ngành/dự án liên quan đến Data Science qua việc học những công nghệ mới nhất về Nghiên cứu và Phân tích Dữ liệu
10. Có thể dễ dàng học lên level của Data Scientist và Data Science Leader/Manager

Các ứng dụng có thể thực hiện sau khóa học:

- Phân tích dữ liệu kinh doanh
- Phân tích dữ liệu nhân sự
- Phân tích dữ liệu điều tra thị trường
- Phân tích dữ liệu khách hàng
- Phân tích dữ liệu tài chính
- Phân tích dữ liệu dòng tiền doanh nghiệp

Các lĩnh vực có thể ứng dụng: phân tích gian lận trong các giao dịch tài chính, thương mại (fraud analysis), hoạch định nguồn lực doanh nghiệp (resource planning), bán lẻ (retail), chứng khoán (stock exchange), vận tải và giao nhận hàng hóa (delivery logistics), y tế và chăm sóc sức khỏe (healthcare), quản lý kho bãi (inventory management), marketing, chăm sóc khách hàng,...

Yêu cầu đầu vào: Tin học (sử dụng máy tính cơ bản), Toán (toán phổ thông), Tiếng Anh (đọc hiểu cơ bản, tương đương bằng A)

ĐỀ CƯƠNG CHI TIẾT

Tuần	Nội dung	Thực hành	Kết quả đạt được
01 Giới thiệu về Data Science và Python	Buổi 01: Nhập môn <ul style="list-style-type: none"> • Tổng quan về Data Science và Data Analytics • Làm quen với lập trình Python 	<ul style="list-style-type: none"> - Cài đặt Anaconda - Cài đặt Jupyter - Cài đặt chương trình soạn thảo mã nguồn 	<ul style="list-style-type: none"> - Hiểu về khái niệm Data Science và các khái niệm cơ bản liên quan - Biết cách cài đặt, viết mã nguồn và thực thi chương trình Python
	Buổi 02: Thực hành Làm quen lập trình Python: <ul style="list-style-type: none"> • Các loại dữ liệu cơ bản • Các phép toán cơ bản • Biến/Hằng • Hàm 	<ul style="list-style-type: none"> - Viết và thực thi chương trình Python đơn giản - Thực hiện các phép toán cơ bản trong Python - Sử dụng biến trong Python - Sử dụng hàm (function) trong Python 	<ul style="list-style-type: none"> - Nắm bắt được các quy tắc cơ bản (syntax) khi viết chương trình Python - Thực hiện được các lệnh cơ bản - Biết cách sử dụng biến - Biết quy tắc viết hàm

02	Buổi 01: Lập trình Python <ul style="list-style-type: none"> Câu lệnh điều kiện Vòng lặp Tư duy thuật toán 	Thực hành ở buổi tiếp theo	<ul style="list-style-type: none"> Hiểu về quy tắc viết câu điều kiện Nắm bắt được các trường hợp sử dụng vòng lặp và quy tắc tạo vòng lặp Cách tạo các hàm để hiện thực thuật toán
	Buổi 02: Thực hành <ul style="list-style-type: none"> Các nội dung ở buổi 01 	<ul style="list-style-type: none"> Luyện tập cách sử dụng vòng lặp, các lệnh điều kiện "if/else" Các vòng lặp for, while Các phép toán thoát, dừng vòng lặp Luyện tập cách xây dựng thuật toán và chuyển thành mã nguồn 	<ul style="list-style-type: none"> Thực hiện thành thạo việc sử dụng câu điều kiện để lựa chọn các trường hợp Thực hiện thành thạo vòng lặp for, while Biết cách thoát vòng lặp sớm
03	Buổi 01: Lập trình Python (tt.) <ul style="list-style-type: none"> List/Array Đọc/ghi dữ liệu Các hàm thống kê đơn giản trong Python 	Thực hành ở buổi tiếp theo	<ul style="list-style-type: none"> Hiểu được các dữ liệu dạng chuỗi trong Python Nắm bắt được các quy tắc đọc/ghi file đơn giản Hiểu được ý nghĩa và cách tính các thống kê cơ bản
	Buổi 02: Thực hành <ul style="list-style-type: none"> Các nội dung ở buổi 01 	<ul style="list-style-type: none"> Thực hiện các phép toán, vòng lặp trên dữ liệu dạng chuỗi Thực hiện ghi dữ liệu lên file và đọc dữ liệu từ file Thực hiện ghi/đọc file csv một cách đơn giản Thực hiện tính các thống kê cơ bản trong Python (mean, median, min, max, range) 	<ul style="list-style-type: none"> Biết cách tổ chức dữ liệu dạng chuỗi Phân biệt được một số kiểu dữ liệu dạng chuỗi cơ bản (list/set/tuple) Nhanh chóng làm quen các hàm thống kê đơn giản
04	Buổi 01: Làm quen thư viện Matplotlib	<ul style="list-style-type: none"> Các quy tắc và cách trình bày biểu đồ trong Matplotlib Cách tạo các biểu đồ con, sắp xếp các biểu đồ, chọn màu Trình bày dữ liệu chứng khoán ở dạng biểu đồ đơn giản 	<ul style="list-style-type: none"> Nhanh chóng làm quen và cảm nhận được khả năng vẽ biểu đồ của Python Nắm rõ được các quy tắc trình bày các thành phần biểu đồ trong Matplotlib
	Buổi 02: Các loại biểu đồ đơn giản	<ul style="list-style-type: none"> Luyện tập cách vẽ các biểu đồ đơn giản: dạng điểm, đường, cột, hình tròn, scatter Thực hiện chỉnh sửa text minh họa trong biểu đồ 	<ul style="list-style-type: none"> Biết cách vẽ các biểu đồ cơ bản trong Python Làm quen cách viết mã nguồn vẽ và trình bày các biểu đồ đơn giản cho phân tích dữ liệu
05	Buổi 01: Xử lý dữ liệu số với Numpy	<ul style="list-style-type: none"> Hiểu được các quy tắc cơ bản khi sử dụng Numpy Luyện tập các xử lý trên dữ liệu dạng mảng 	<ul style="list-style-type: none"> Biết cách thực hiện các xử lý trên dữ liệu số phục vụ cho thống kê mô tả Hiểu được dữ liệu dạng mảng Biết cách sử dụng Numpy để thực hiện các thống kê

	Buổi 02: Xử lý dữ liệu text và thời gian	<ul style="list-style-type: none"> - Thực hiện các xử lý cắt, ghép, chỉnh sửa, trên chuỗi - Thực hiện convert dữ liệu thời gian sang text và ngược lại - Thực hiện các phép tính thời gian 	<ul style="list-style-type: none"> - Nắm bắt được các hàm xử lý dữ liệu dạng text - Biết cách xử lý dữ liệu dạng thời gian - Biết cách chuyển đổi dữ liệu thời gian từ text
06 Phân tích Dữ liệu Đơn giản	Buổi 01: Phân tích mô tả (descriptive analytics)	<ul style="list-style-type: none"> - Thực hành các phân tích thống kê cơ bản phục vụ cho phân tích mô tả - Thực hành phân tích dữ liệu bán lẻ - Thảo luận các kịch bản kinh doanh dựa trên phân tích dữ liệu 	<ul style="list-style-type: none"> - Phân biệt được các loại phân tích thống kê cơ bản (mô tả, dự đoán /dự báo và đề nghị) - Hiểu được phân tích mô tả là gì, và các thông số phổ biến của PTMT - Biết cách trình bày kết quả phân tích thống kê mô tả bằng hình ảnh
	Buổi 02: Báo cáo giữa kỳ	- Học viên chọn một loại dữ liệu liên quan trực tiếp đến lĩnh vực của mình và thực hiện đánh giá dataset đó, đồng thời phân tích mô tả	- Luyện tập các làm việc nhóm để nắm bắt được các yêu cầu, nhận định được khó khăn, thách thức khi tiến hành một dự án phân tích dữ liệu
07 Thư viện Pandas	Buổi 01: Làm quen và sử dụng thư viện pandas	<ul style="list-style-type: none"> - Các loại mảng cơ bản trong Pandas: Series và DataFrame - Đọc dữ liệu từ file .csv, excel vào khung dữ liệu (dataframe) bằng thư viện pandas - Các xử lý cơ bản trên Series và DataFrame trong pandas - Phân tích dữ liệu bán hàng và vẽ biểu đồ trình bày kết quả với pandas 	<ul style="list-style-type: none"> - Hiểu được các kiến thức cơ bản để sử dụng thư viện pandas - Biết cách phân tách, ghép nối các dataframe phục vụ cho việc phân tích dữ liệu
	Buổi 02: Xử lý dữ liệu nâng cao và thực hiện làm sạch dữ liệu với pandas	<ul style="list-style-type: none"> - Thực hành làm sạch dữ liệu với pandas và numpy - Thực hiện chia nhỏ và load các tập dữ liệu có kích thước lớn bằng chunk 	<ul style="list-style-type: none"> - Hiểu được các ứng dụng từ cơ bản đến nâng cao của thư viện pandas trong việc xử lý, làm sạch, chuyển đổi... dữ liệu - Biết cách xử lý khi dữ liệu không đầy đủ (missing data) bằng pandas - Biết cách sử dụng pandas để load và xử lý những tập dữ liệu lớn
08 Phân tích Dữ liệu Nâng cao	Buổi 01: Phân tích dự đoán (predictive analytics) và Phân tích đề xuất (prescriptive analytics)	<ul style="list-style-type: none"> - Hồi quy tuyến tính (linear regression) và hồi quy logistic (logistic regression) - Xây dựng chương trình dự đoán doanh thu, giá nhà, tình trạng xe hơi cũ dựa trên dữ liệu lịch sử 	<ul style="list-style-type: none"> - Hiểu được khái niệm và các bước cần thực hiện trong phân tích dự báo - Bước đầu được làm quen với AI (trí tuệ nhân tạo) và ML (machine learning)

	Buổi 02: Phân tích dữ liệu thăm dò (EDA) cơ bản	<ul style="list-style-type: none"> - Thực hiện use-case về phân tích EDA cho doanh nghiệp bán lẻ - Làm quen với việc phân tích tập dữ liệu lớn (>33 triệu dòng) 	<ul style="list-style-type: none"> - Hiểu được khái niệm EDA và các bước cơ bản để thực hiện EDA - Biết cách thực hiện các phân tích cơ bản với lượng dữ liệu lớn
09 Vẽ biểu đồ với Python - Phần II	Buổi 01: Trình bày biểu đồ nâng cao sử dụng Matplotlib	<ul style="list-style-type: none"> - Các loại biểu đồ phức tạp: biểu đồ đa trục với scale khác nhau, biểu đồ dạng bọt, donut và nested-donut... - Trình bày biểu đồ dữ liệu bán hàng, luyện tập cách tạo biểu đồ phức hợp, lồng ghép các loại biểu đồ 	<ul style="list-style-type: none"> - Có khả năng xây dựng các biểu đồ phức hợp, phức tạp để trình bày các kết quả DA đa dạng - Nắm bắt một số biểu đồ có tính thống kê để thực hiện thăm dò dữ liệu
	Buổi 02: Trình bày biểu đồ thống kê với thư viện Seaborn	<ul style="list-style-type: none"> - Vẽ và trang trí các biểu đồ bằng thư viện Seaborn - Vẽ nhanh các biểu đồ thể hiện sự tương quan với PairGrid và PairPlot - Thực hiện một số phân tích thăm dò trên dữ liệu với Seaborn 	<ul style="list-style-type: none"> - Hiểu được các quy tắc cơ bản khi sử dụng thư viện Seaborn - Biết cách trình bày và trang trí biểu đồ sao cho bắt mắt hơn bằng Seaborn - Biết cách áp dụng Seaborn cho một số phân tích thăm dò
10 Dashboard	Buổi 01: Làm quen với dashboard	<ul style="list-style-type: none"> - Làm quen với ipywidgets - Thể hiện các biểu đồ trên một dashboard đơn giản để báo cáo kết quả phân tích dữ liệu bán hàng 	<ul style="list-style-type: none"> - Biết cách kết hợp các công cụ/kỹ thuật đã học để xây dựng một dash-board có tính tương tác nhằm trình bày kết quả phân tích dữ liệu
	Buổi 02: Tạo Figures tương tác để xây dựng dashboard	<ul style="list-style-type: none"> - Làm quen thư viện Plotly - Trình bày biểu đồ trên dashboard cơ bản (giao diện web) 	<ul style="list-style-type: none"> - Biết cách vẽ biểu đồ trên dashboard sử dụng các thư viện Dash và Plotly web)
11 SQL	Buổi 01: Truy vấn SQL trong Python	<ul style="list-style-type: none"> - Học cách kết nối cơ sở dữ liệu đơn giản (sqlite) trong Python - Các mệnh đề SQL cơ bản 	<ul style="list-style-type: none"> - Biết cách sử dụng SQL để truy vấn, làm sạch và phân tích dữ liệu - Biết cách sử dụng các thư viện SQL trong Python (như sqlite)
	Buổi 02: Sử Dụng SQL cho EDA	<ul style="list-style-type: none"> - Truy vấn dữ liệu từ cơ sở dữ liệu bán hàng bằng SQL - Thực hiện phân tích trong Python từ dữ liệu đã truy vấn và trình bày kết quả bằng biểu đồ 	<ul style="list-style-type: none"> - Biết cách tạo dựng, sử dụng các truy vấn SQL phức tạp hoặc kết hợp SQL với pandas để thực hiện phân tích dữ liệu - Hiểu về NoSQL
12	Buổi 01: Lập trình Hướng đối tượng (OOP) và demo Blockchain	<ul style="list-style-type: none"> - Một số quy tắc lập trình OOP trong Python - Demo Blockchain 	<ul style="list-style-type: none"> - Biết cách xây dựng một chương trình OOP đơn giản - Hiểu về cách hoạt động cơ bản của blockchain

	<i>Buổi 02: Báo cáo cuối kỳ</i>	- Trình bày và thảo luận về phương pháp, kỹ thuật, yêu cầu, thách thức, kết quả ... của một dự án phân tích dữ liệu	- Các học viên hoàn thành báo cáo cuối kỳ sẽ được cấp chứng chỉ của khóa học
--	--	---	--